	Type	L #	Hits	Search Text	DBs	Time Stamp	Comments	Error Definition	Errors
1	BRS	L1	2896	codon same optimiz$5	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/02/28 17:57			0
2	BRS	L2	12706	(human adj protein) or (factor adj VIII) or (factor adj iX)	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/02/28 17:58			0
3	BRS	L3	6	1 same 2	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/02/28 18:05			0
4	BRS	L4	3899	high adj express$3	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/02/28 18:06			0
5	BRS	L5	23	2 same 4	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/02/28 18:07			0
6	BRS	L6	1239	codon same (common or preferred)	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/02/28 18:07			0
7	BRS	L7	0	5 same 6	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/02/28 18:07			0
8	BRS	L8	3	5 same codon	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/02/28 18:08			0
9	BRS	L9	27	(common adj codon)	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/02/28 18:10			0
10	BRS	L10	2	2 same 9	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/02/28 18:12			0
11	BRS	L11	1	(non$1common adj codon)	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/02/28 18:13			0
12	BRS	L12	228	(primary or secondary) adj mammalian adj cell	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/02/28 18:14			0
13	BRS	L13	879	(nucleic adj acid) same 1	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/02/28 18:15			0
14	BRS	L14	6	2 same 3	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/02/28 18:15			0
15	BRS	L15	0	12 same 14	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/02/28 18:16			0
16	BRS	L16	2920	episome	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/02/28 18:16			0
17	BRS	L17	0	14 same 16	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/02/28 18:16			0

	Type	L #	Hits	Search Text	DBs	Time Stamp	Comments	Error Definition	Errors
18	BRS	L18	54	miller adj allan.in.	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/02/28 18:16			0
19	BRS	L19	39	treco adj douglas.in.	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/02/28 18:17			0
20	BRS	L20	1	seldon adj richard.in.	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/02/28 18:17			0
21	BRS	L21	9	(19 or 18 or 20) and 1	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/02/28 18:18			0

FILE 'HOME' ENTERED AT 19:07:54 ON 28 FEB 2003

=> file medline caplus biosis embase scisearch agricola
COST IN U.S. DOLLARS SINCE FILE TOTAL
ENTRY SESSION
FULL ESTIMATED COST 0.21 0.21

FILE 'MEDLINE' ENTERED AT 19:08:18 ON 28 FEB 2003

FILE 'CAPLUS' ENTERED AT 19:08:18 ON 28 FEB 2003
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.

FILE 'BIOSIS' ENTERED AT 19:08:18 ON 28 FEB 2003

FILE 'EMBASE' ENTERED AT 19:08:18 ON 28 FEB 2003

FILE 'SCISEARCH' ENTERED AT 19:08:18 ON 28 FEB 2003

FILE 'AGRICOLA' ENTERED AT 19:08:18 ON 28 FEB 2003

=> s codon (p) optimiz?
L1 1557 CODON (P) OPTIMIZ?

=> s (human protein) or (factor VIII) or (factor IX)
3 FILES SEARCHED...
L2 87456 (HUMAN PROTEIN) OR (FACTOR VIII) OR (FACTOR IX)

=> s l1 (p) l2
L3 13 L1 (P) L2

=> duplicate remove l3
DUPLICATE PREFERENCE IS 'MEDLINE, CAPLUS, BIOSIS, EMBASE, SCISEARCH'
KEEP DUPLICATES FROM MORE THAN ONE FILE? Y/(N):n
PROCESSING COMPLETED FOR L3
L4 5 DUPLICATE REMOVE L3 (8 DUPLICATES REMOVED)

=> d l4 1-5 ibib abs

L4 ANSWER 1 OF 5 CAPLUS COPYRIGHT 2003 ACS
ACCESSION NUMBER: 2002:637845 CAPLUS
DOCUMENT NUMBER: 137:180783
TITLE: Synthetic genes with optimized codon usage for recombinant protein expression in mammals
INVENTOR(S): Seldon, Richard F.; Miller, Allan M.; Treco, Douglas S.
PATENT ASSIGNEE(S): Transkaryotic Therapies, Inc., USA
SOURCE: PCT Int. Appl., 115 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2002064799	A2	20020822	WO 2001-US42655	20011011

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG

PRIORITY APPLN. INFO.: US 1999-407605 A1 19990929
US 2000-686497 A1 20001011

AB The present invention is directed to a synthetic nucleic acid sequence which encodes a protein wherein at least one non-common codon or less-common codon is replaced by a common codon. The synthetic nucleic acid sequence can include a continuous stretch of at least 90 codons all of which are common codons. Synthetic genes that have codon usages typical of highly expressed mammalian genes are prepd. for use in manuf. of the proteins using mammalian cell hosts. Recombinant expression of human Factor VIII and B-domain-deleted-FVIII, human Factor IX, and human .alpha.-galactosidase, in human fibroblast cells, is described.

L4 ANSWER 2 OF 5 MEDLINE DUPLICATE 1
ACCESSION NUMBER: 2000494820 MEDLINE
DOCUMENT NUMBER: 20442636 PubMed ID: 10985959
TITLE: Fusion protein vectors to increase protein production and evaluate the immunogenicity of genetic vaccines.
AUTHOR: Wu L; Barry M A

CORPORATE SOURCE: Center for Cell and Gene Therapy, Baylor College of Medicine, Houston, Texas, 77030, USA.
CONTRACT NUMBER: AI042588 (NIAID)
AI36211 (NIAID)
SOURCE: MOLECULAR THERAPY, (2000 Sep) 2 (3) 288-97.
Journal code: 100890581. ISSN: 1525-0016.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals; AIDS
ENTRY MONTH: 200010
ENTRY DATE: Entered STN: 20001027
Last Updated on STN: 20001027
Entered Medline: 20001013

AB Genetic immunization is a method for vaccination and laboratory antibody production where antigen-expressing plasmids are introduced into animals to elicit immune responses. Although genetic immunization works well for many antigens, problems can arise with protein sequences that (i) are toxic to host cells, (ii) are difficult to translate by mammalian cells, or (iii) evade immune presentation. We demonstrate here the ability to increase protein production and antigen secretion by the simple method of fusing poorly expressed sequences to well-expressed heterologous proteins. Proof-of-principle is demonstrated here using the poorly translated HIV-1 envelope whose protein production is rescued by fusing this antigen to the carboxy-termini of two well-expressed proteins: the cytoplasmic green fluorescent protein and the secreted ***human*** ***protein*** a1-antitrypsin. This approach represents a simple and substantially less expensive method to increase protein and antigen production than ***codon*** - ***optimization*** strategies. It may therefore be more useful than whole gene ***codon*** replacement to enable inexpensive laboratory antibody production of poorly expressed antigens and for large-scale genomic protein or antigen screening efforts. Finally, we demonstrate a second benefit of this antigen fusion strategy in which the test antigen is "sandwiched" between two positive control antigens. By this approach, we demonstrate the intrinsic lack of immunogenicity of HIV-1 envelope under conditions when robust antibody responses are generated against its fusion protein partners, but not against this evasive antigen. These fusion protein vectors therefore represent a simple approach to not only increase antigen production, but also assess antigen production and immunogenicity in vivo.

L4 ANSWER 3 OF 5 MEDLINE DUPLICATE 2
ACCESSION NUMBER: 1998192613 MEDLINE
DOCUMENT NUMBER: 98192613 PubMed ID: 9525926
TITLE: Improved fluorescence and dual color detection with enhanced blue and green variants of the green fluorescent

protein.
AUTHOR: Yang T T; Sinai P; Green G; Kitts P A; Chen Y T; Lybarger L; Chervenak R; Patterson G H; Piston D W; Kain S R
CORPORATE SOURCE: Cell Biology Group, Clontech Laboratories, Inc., Palo Alto, California 94303, USA.
SOURCE: JOURNAL OF BIOLOGICAL CHEMISTRY, (1998 Apr 3) 273 (14) 8212-6.
Journal code: 2985121R. ISSN: 0021-9258.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 199805
ENTRY DATE: Entered STN: 19980514
Last Updated on STN: 19980514
Entered Medline: 19980507
AB The green fluorescent protein (GFP) from the jellyfish Aequorea victoria is a versatile reporter protein for monitoring gene expression and protein localization in a variety of systems. Applications using GFP reporters have expanded greatly due to the availability of mutants with altered spectral properties, including several blue emission variants, all of which contain the single point mutation Tyr-66 to His in the chromophore region of the protein. However, previously described "BFP" reporters have limited utility, primarily due to relatively dim fluorescence and low expression levels attained in higher eukaryotes with such variants. To improve upon these qualities, we have combined a blue emission mutant of GFP containing four point mutations (Phe-64 to Leu, Ser-65 to Thr, Tyr-66 to His, and Tyr-145 to Phe) with a synthetic gene sequence containing ***codons*** preferentially found in highly expressed ***human*** ***proteins*** . These mutations were chosen to ***optimize*** expression of properly folded fluorescent protein in mammalian cells cultured at 37 degreesC and to maximize signal intensity. The combination of improved fluorescence and higher expression levels yield an enhanced blue fluorescent protein that provides greater sensitivity and is suitable for dual color detection with green-emitting fluorophores.

L4 ANSWER 4 OF 5 CAPLUS COPYRIGHT 2003 ACS DUPLICATE 3
ACCESSION NUMBER: 1998:88435 CAPLUS
DOCUMENT NUMBER: 128:166966
TITLE: An Integrated Sequence-Structure Database incorporating matching mRNA sequence, amino acid sequence and protein three-dimensional structure data
AUTHOR(S): Adzhubei, Ivan A.; Adzhubei, Alexei A.; Neidle, Stephen
CORPORATE SOURCE: CRC Biomolecular Structure Unit, The Institute of Cancer Research, Surrey, SM2 5NG, UK

SOURCE: Nucleic Acids Research (1998), 26(1), 327-331
CODEN: NARHAD; ISSN: 0305-1048
PUBLISHER: Oxford University Press
DOCUMENT TYPE: Journal
LANGUAGE: English

AB We have constructed a non-homologous database, termed the Integrated Sequence-Structure Database (ISSD) which comprises the coding sequences of genes, amino acid sequences of the corresponding proteins, their secondary structure and .vphi.,.psi. angles assignments, and polypeptide backbone coordinates. Each protein entry in the database holds the alignment of nucleotide sequence, amino acid sequence and the PDB three-dimensional structure data. The nucleotide and amino acid sequences for each entry are selected on the basis of exact matches of the source organism and cell environment. The current version 1.0 of ISSD is available on the WWW at http://www.protein.bio.msu.su/issd/ and includes 107 non-homologous mammalian proteins, of which 80 are ***human*** ***proteins*** . The database has been used by us for the anal. of synonymous ***codon*** usage patterns in mRNA sequences showing their correlation with the three-dimensional structure features in the encoded proteins. Possible ISSD applications include ***optimization*** of protein expression, improvement of the protein structure prediction accuracy, and anal. of evolutionary aspects of the nucleotide sequence-protein structure relationship.

L4 ANSWER 5 OF 5 MEDLINE DUPLICATE 4
ACCESSION NUMBER: 92119062 MEDLINE
DOCUMENT NUMBER: 92119062 PubMed ID: 1768766
TITLE: Genetics and molecular biology of haemophilias A and B.
AUTHOR: Green P M; Montandon A J; Bentley D R; Giannelli F
CORPORATE SOURCE: Division of Medical and Molecular Genetics, United Medical School of Guy's Hospital, London Bridge, UK.
SOURCE: BLOOD COAGULATION AND FIBRINOLYSIS, (1991 Aug) 2 (4) 539-65. Ref: 178
Journal code: 9102551. ISSN: 0957-5235.
PUB. COUNTRY: ENGLAND: United Kingdom
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
General Review; (REVIEW)
(REVIEW, ACADEMIC)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 199202
ENTRY DATE: Entered STN: 19920315
Last Updated on STN: 19990129
Entered Medline: 19920227

AB The development of rapid procedures for the characterization of mutations is advancing the knowledge of the molecular biology of the haemophilias

and transforming the strategies for the diagnoses required for genetic counselling. In haemophilia B more than 300 mutants have been fully characterized. These comprise complete and partial deletions, rare insertions, and 'point' mutations. The latter may impair transcription (promoter mutations), RNA processing (splicing mutations) and translation (frameshifts and stop ***codons***) or cause single amino acid (aa) changes. Eighty-four residues are involved in the 105 presumed detrimental aa substitutions reported so far and these are usually conserved in the ***factor*** ***IX*** homologues (factors VII, X and protein C) and/or the ***factor*** ***IX*** of different mammalian species. There are clear correlations between the mutation and clinical features. In addition mutations causing gross physical or functional loss of coding information appear to predispose to the development of antibodies against therapeutic ***factor*** ***IX*** . Hotspots of mutations have been identified and are usually associated with CpG sequences. In haemophilia A the size and complexity of the ***factor*** ***VIII*** gene has hindered the analysis of mutants. Most of the studies published so far have analysed only a small fraction of the essential region of the ***factor*** ***VIII*** gene and this led to the repeated observation of specific types of mutation. The recent development of a rapid method to analyse RNA splicing and the whole coding region of the ***factor*** ***VIII*** gene should unblock this situation. With regard to genetic counselling, the direct detection of gene defects has increased the proportion of haemophilia B families that can be helped from 60% to virtually 100% and similar expectations may now be formulated for haemophilia A. In the UK a national database of haemophilia B mutations is being constructed to ***optimize*** genetic counselling. This should offer a model for a similar development in haemophilia A.

=> s non-common codon
L5 1 NON-COMMON CODON

=> d l5 1 ibib abs

L5 ANSWER 1 OF 1 CAPLUS COPYRIGHT 2003 ACS
ACCESSION NUMBER: 2002:637845 CAPLUS
DOCUMENT NUMBER: 137:180783
TITLE: Synthetic genes with optimized codon usage for recombinant protein expression in mammals
INVENTOR(S): Seldon, Richard F.; Miller, Allan M.; Treco, Douglas S.
PATENT ASSIGNEE(S): Transkaryotic Therapies, Inc., USA
SOURCE: PCT Int. Appl., 115 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent

LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2002064799	A2	20020822	WO 2001-US42655	20011011

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG
PRIORITY APPLN. INFO.: US 1999-407605 A1 19990929
US 2000-686497 A1 20001011

AB The present invention is directed to a synthetic nucleic acid sequence which encodes a protein wherein at least one ***non*** - ***common*** ***codon*** or less-common codon is replaced by a common codon. The synthetic nucleic acid sequence can include a continuous stretch of at least 90 codons all of which are common codons. Synthetic genes that have codon usages typical of highly expressed mammalian genes are prepd. for use in manuf. of the proteins using mammalian cell hosts. Recombinant expression of human Factor VIII and B-domain-deleted-FVIII, human Factor IX, and human .alpha.-galactosidase, in human fibroblast cells, is described.

=> d his

(FILE 'HOME' ENTERED AT 19:07:54 ON 28 FEB 2003)

FILE 'MEDLINE, CAPLUS, BIOSIS, EMBASE, SCISEARCH, AGRICOLA' ENTERED AT
19:08:18 ON 28 FEB 2003
L1 1557 S CODON (P) OPTIMIZ?
L2 87456 S (HUMAN PROTEIN) OR (FACTOR VIII) OR (FACTOR IX)
L3 13 S L1 (P) L2
L4 5 DUPLICATE REMOVE L3 (8 DUPLICATES REMOVED)
L5 1 S NON-COMMON CODON

=> s (primary or secondary) (w) (mammalian cell)
4 FILES SEARCHED...
L6 93 (PRIMARY OR SECONDARY) (W) (MAMMALIAN CELL)

=> s l6 (p) (nucleic acid)
L7 2 L6 (P) (NUCLEIC ACID)

=> s l7 (p) codon
L8 0 L7 (P) CODON

=> s episome
L9 40436 EPISOME

=> s l7 (p) l9
L10 0 L7 (P) L9

=> d his

(FILE 'HOME' ENTERED AT 19:07:54 ON 28 FEB 2003)

FILE 'MEDLINE, CAPLUS, BIOSIS, EMBASE, SCISEARCH, AGRICOLA' ENTERED AT
19:08:18 ON 28 FEB 2003
L1 1557 S CODON (P) OPTIMIZ?
L2 87456 S (HUMAN PROTEIN) OR (FACTOR VIII) OR (FACTOR IX)
L3 13 S L1 (P) L2
L4 5 DUPLICATE REMOVE L3 (8 DUPLICATES REMOVED)
L5 1 S NON-COMMON CODON
L6 93 S (PRIMARY OR SECONDARY) (W) (MAMMALIAN CELL)
L7 2 S L6 (P) (NUCLEIC ACID)
L8 0 S L7 (P) CODON
L9 40436 S EPISOME
L10 0 S L7 (P) L9

=> log y
COST IN U.S. DOLLARS SINCE FILE TOTAL
ENTRY SESSION
FULL ESTIMATED COST 48.29 48.50

DISCOUNT AMOUNTS (FOR QUALIFYING ACCOUNTS) SINCE FILE TOTAL
ENTRY SESSION
CA SUBSCRIBER PRICE -1.95 -1.95

STN INTERNATIONAL LOGOFF AT 19:13:54 ON 28 FEB 2003